

THE RELATIVE EFFICIENCIES OF HAPLOTYPE AND GENOTYPE DATA FOR FINE-SCALE LD MAPPING

David Balding¹, Andrew Morris², John Whittaker¹

¹Department of Epidemiology and Public Health, Imperial College London, London; ²Wellcome Trust Centre for Human Genetics, Oxford, UK

A common approach to the analysis of multi-locus genotype data for linkage disequilibrium (LD) fine mapping is a two-stage strategy of first inferring the haplotype pairs underlying the genotypes and then implementing a haplotype-based analysis. This is unsatisfactory for a number of reasons. The uncertainty arising in the haplotype inferral stage is often ignored in the second stage. Worse, errors in haplotype inferral are often in the direction of maximising apparent LD, leading to spurious precision of subsequent analyses. Because true haplotypes are more informative for fine mapping than the corresponding genotypes, users may erroneously believe that inferred haplotype pairs are also more informative than the original genotypes. We present the results of a simulation study that indicate that genotype data, appropriately analysed, often provide almost as much information for LD mapping as do true haplotype pairs, and generally provide more information than do inferred haplotypes. To perform the study, we used our coalescent-based MCMC software COLDMAP (COalescent-based LD MAPping). Full results will appear in *American Journal of Human Genetics*, probably in May 2004.

Please indicate if this abstract is being submitting for the poster session only, for a possible presentation, or for both.